

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
18 April 2002 (18.04.2002)

PCT

(10) International Publication Number
WO 02/31147 A2

(51) International Patent Classification⁷: C12N 15/12, C07K 14/705

(21) International Application Number: PCT/US01/31113

(22) International Filing Date: 4 October 2001 (04.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/239,629 10 October 2000 (10.10.2000) US

(71) Applicant: LEXICON GENETICS INCORPORATED
[US/US]; 4000 Research Forest Drive, The Woodlands, TX 77381 (US).

(72) Inventors: HU, Yi; 333 Holly Creek Ct. #203, The Woodlands, TX 77381 (US). NEPOMNICHY, Boris; 905 Cypress Station Dr. #1515, Houston, TX 77909 (US).

(74) Agents: ISHIMOTO, Lance, K. et al.; Lexicon Genetics Incorporated, 4000 Research Forest Drive, The Woodlands, TX 77381 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL HUMAN TRANSPORTER PROTEINS AND POLYNUCLEOTIDES ENCODING THE SAME

(57) Abstract: Novel human polynucleotide and polypeptide sequences are disclosed that can be used in therapeutic, diagnostic, and pharmacogenomic applications.

WO 02/31147 A2

NOVEL HUMAN TRANSPORTER PROTEINS AND
POLYNUCLEOTIDES ENCODING THE SAME

The present application claims the benefit of U.S.
Provisional Application Number 60/239,629, which was filed on
5 October 10, 2000, and is herein incorporated by reference in its
entirety.

1. INTRODUCTION

The present invention relates to the discovery,
identification, and characterization of novel human
10 polynucleotides encoding proteins that share sequence similarity
with mammalian transporter proteins. The invention encompasses
the described polynucleotides, host cell expression systems, the
encoded proteins, fusion proteins, polypeptides and peptides,
antibodies to the encoded proteins and peptides, and genetically
15 engineered animals that either lack or over express the disclosed
genes, antagonists and agonists of the proteins, and other
compounds that modulate the expression or activity of the
proteins encoded by the disclosed genes that can be used for
diagnosis, drug screening, clinical trial monitoring, the
20 treatment of diseases and disorders, and cosmetic or
nutriceutical applications.

2. BACKGROUND OF THE INVENTION

Transporter proteins are integral membrane proteins that
mediate or facilitate the passage of materials across the lipid
25 bilayer. Given that the transport of materials across the
membrane can play an important physiological role, transporter
proteins are good drug targets. Additionally, one of the
mechanisms of drug resistance involves diseased cells using
cellular transporter systems to export chemotherapeutic agents
30 from the cell. Such mechanisms are particularly relevant to
cells manifesting resistance to a multiplicity of drugs.

3. SUMMARY OF THE INVENTION

The present invention relates to the discovery, identification, and characterization of nucleotides that encode novel human proteins, and the corresponding amino acid sequences of these proteins. The novel human proteins (NHPs) described for the first time herein share structural similarity with mammalian ATP-binding cassette (ABC) transporters and multidrug resistance transporters.

The novel human nucleic acid sequences described herein, encode alternative proteins/open reading frames (ORFs) of 1,642 and 1,594 amino acids in length (see respectively SEQ ID NOS: 2 and 4).

The invention also encompasses agonists and antagonists of the described NHPs, including small molecules, large molecules, mutant NHPs, or portions thereof, that compete with native NHP, peptides, and antibodies, as well as nucleotide sequences that can be used to inhibit the expression of the described NHPs (e.g., antisense and ribozyme molecules, and open reading frame or regulatory sequence replacement constructs) or to enhance the expression of the described NHPs (e.g., expression constructs that place the described polynucleotide under the control of a strong promoter system), and transgenic animals that express a NHP sequence, or "knock-outs" (which can be conditional) that do not express a functional NHP. Knock-out mice can be produced in several ways, one of which involves the use of mouse embryonic stem cells ("ES cells") lines that contain gene trap mutations in a murine homolog of at least one of the described NHPs. When the unique NHP sequences described in SEQ ID NOS:1-5 are "knocked-out" they provide a method of identifying phenotypic expression of the particular gene as well as a method of assigning function to previously unknown genes. In addition, animals in which the unique NHP sequences described in SEQ ID NOS:1-5 are "knocked-out" provide a unique source in which to elicit antibodies to

homologous and orthologous proteins that would have been previously viewed by the immune system as "self" and therefore would have failed to elicit significant antibody responses. To these ends, gene trapped knockout ES cells have been generated in
5 murine homologs of the described NHPs.

Additionally, the unique NHP sequences described in SEQ ID NOS:1-5 are useful for the identification of protein coding sequence and mapping a unique gene to a particular chromosome (in this case, human chromosome 17, see GENBANK accession no.
10 AC005495). These sequences identify biologically verified exon splice junctions as opposed to splice junctions that may have been bioinformatically predicted from genomic sequence alone. The sequences of the present invention are also useful as additional DNA markers for restriction fragment length
15 polymorphism (RFLP) analysis, and in forensic biology.

Further, the present invention also relates to processes for identifying compounds that modulate, *i.e.*, act as agonists or antagonists, of NHP expression and/or NHP activity that utilize purified preparations of the described NHPs and/or NHP product,
20 or cells expressing the same. Such compounds can be used as therapeutic agents for the treatment of any of a wide variety of symptoms associated with biological disorders or imbalances.

4. DESCRIPTION OF THE SEQUENCE LISTING AND FIGURES

25 The Sequence Listing provides the sequences of the NHP ORFs encoding the described NHP amino acid sequences. SEQ ID NO:5 shows a NHP ORF and flanking regions.

5. DETAILED DESCRIPTION OF THE INVENTION

30 The NHPs described for the first time herein are novel proteins that may be expressed in, *inter alia*, human cell lines, fetal brain, brain, pituitary, lymph node, kidney, fetal liver,

liver, testis, thyroid, adrenal gland, fetal lung, and fetal kidney cells.

The present invention encompasses the nucleotides presented in the Sequence Listing, host cells expressing such nucleotides, 5 the expression products of such nucleotides, and: (a) nucleotides that encode mammalian homologs of the described genes, including the specifically described NHPs, and the NHP products; (b) nucleotides that encode one or more portions of the NHPs that correspond to functional domains, and the polypeptide products 10 specified by such nucleotide sequences, including but not limited to the novel regions of any active domain(s); (c) isolated nucleotides that encode mutant versions, engineered or naturally occurring, of the described NHPs in which all or a part of at least one domain is deleted or altered, and the polypeptide 15 products specified by such nucleotide sequences, including but not limited to soluble proteins and peptides in which all or a portion of the signal (or hydrophobic transmembrane) sequence is deleted; (d) nucleotides that encode chimeric fusion proteins containing all or a portion of a coding region of an NHP, or one 20 of its domains (e.g., a receptor or ligand binding domain, accessory protein/self-association domain, etc.) fused to another peptide or polypeptide; or (e) therapeutic or diagnostic derivatives of the described polynucleotides such as oligonucleotides, antisense polynucleotides, ribozymes, dsRNA, or 25 gene therapy constructs comprising a sequence first disclosed in the Sequence Listing.

As discussed above, the present invention includes: (a) the human DNA sequences presented in the Sequence Listing (and vectors comprising the same) and additionally contemplates any 30 nucleotide sequence encoding a contiguous NHP open reading frame (ORF) that hybridizes to a complement of a DNA sequence presented in the Sequence Listing under highly stringent conditions, e.g., hybridization to filter-bound DNA in 0.5 M NaHPO₄, 7% sodium

dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel et al., eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York, at p. 2.10.3) and encodes a functionally equivalent expression product. Additionally contemplated are any nucleotide sequences that hybridize to the complement of a DNA sequence that encodes and expresses an amino acid sequence presented in the Sequence Listing under moderately stringent conditions, e.g., washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel et al., 1989, *supra*), yet still encodes a functionally equivalent NHP product. Functional equivalents of a NHP include naturally occurring NHPs present in other species and mutant NHPs whether naturally occurring or engineered (by site directed mutagenesis, gene shuffling, directed evolution as described in, for example, U.S. Patent Nos. 5,837,458 and 5,723,323 both of which are herein incorporated by reference in their entirety). The invention also includes degenerate nucleic acid variants of the disclosed NHP polynucleotide sequences.

Additionally contemplated are polynucleotides encoding NHP ORFs, or their functional equivalents, encoded by polynucleotide sequences that are about 99, 95, 90, or about 85 percent similar or identical to corresponding regions of the nucleotide sequences of the Sequence Listing (as measured by BLAST sequence comparison analysis using, for example, the GCG sequence analysis package using standard default settings).

The invention also includes nucleic acid molecules, preferably DNA molecules, that hybridize to, and are therefore the complements of, the described NHP gene nucleotide sequences. Such hybridization conditions may be highly stringent or less highly stringent, as described above. In instances where the nucleic acid molecules are deoxyoligonucleotides ("DNA oligos"), such molecules are generally about 16 to about 100 bases long, or

about 20 to about 80, or about 34 to about 45 bases long, or any variation or combination of sizes represented therein that incorporate a contiguous region of sequence first disclosed in the Sequence Listing. Such oligonucleotides can be used in
5 conjunction with the polymerase chain reaction (PCR) to screen libraries, isolate clones, and prepare cloning and sequencing templates, etc.

Alternatively, such NHP oligonucleotides can be used as hybridization probes for screening libraries, and assessing gene
10 expression patterns (particularly using a micro array or high-throughput "chip" format). Additionally, a series of the described NHP oligonucleotide sequences, or the complements thereof, can be used to represent all or a portion of the described NHP sequences. An oligonucleotide or polynucleotide
15 sequence first disclosed in at least a portion of one or more of the sequences of SEQ ID NOS: 1-5 can be used as a hybridization probe in conjunction with a solid support matrix/substrate (resins, beads, membranes, plastics, polymers, metal or metallized substrates, crystalline or polycrystalline substrates,
20 etc.). Of particular note are spatially addressable arrays (i.e., gene chips, microtiter plates, etc.) of oligonucleotides and polynucleotides, or corresponding oligopeptides and polypeptides, wherein at least one of the biopolymers present on the spatially addressable array comprises an oligonucleotide or
25 polynucleotide sequence first disclosed in at least one of the sequences of SEQ ID NOS: 1-5, or an amino acid sequence encoded thereby. Methods for attaching biopolymers to, or synthesizing biopolymers on, solid support matrices, and conducting binding studies thereon are disclosed in, *inter alia*, U.S. Patent Nos.
30 5,700,637, 5,556,752, 5,744,305, 4,631,211, 5,445,934, 5,252,743, 4,713,326, 5,424,186, and 4,689,405 the disclosures of which are herein incorporated by reference in their entirety.

Addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-5 can be used to identify and characterize the temporal and tissue specific expression of a gene. These addressable arrays incorporate oligonucleotide sequences of sufficient length to confer the required specificity, yet be within the limitations of the production technology. The length of these probes is within a range of between about 8 to about 2000 nucleotides. Preferably the probes consist of 60 nucleotides and more preferably 25 nucleotides from the sequences first disclosed in SEQ ID NOS:1-5.

For example, a series of the described oligonucleotide sequences, or the complements thereof, can be used in chip format to represent all or a portion of the described sequences. The oligonucleotides, typically between about 16 to about 40 (or any whole number within the stated range) nucleotides in length can partially overlap each other and/or the sequence may be represented using oligonucleotides that do not overlap. Accordingly, the described polynucleotide sequences shall typically comprise at least about two or three distinct oligonucleotide sequences of at least about 8 nucleotides in length that are each first disclosed in the described Sequence Listing. Such oligonucleotide sequences can begin at any nucleotide present within a sequence in the Sequence Listing and proceed in either a sense (5'-to-3') orientation vis-a-vis the described sequence or in an antisense orientation.

Microarray-based analysis allows the discovery of broad patterns of genetic activity, providing new understanding of gene functions and generating novel and unexpected insight into transcriptional processes and biological mechanisms. The use of addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-5 provides detailed information about transcriptional changes involved in a specific pathway, potentially leading to

the identification of novel components or gene functions that manifest themselves as novel phenotypes.

Probes consisting of sequences first disclosed in SEQ ID NOS:1-5 can also be used in the identification, selection and validation of novel molecular targets for drug discovery. The use of these unique sequences permits the direct confirmation of drug targets and recognition of drug dependent changes in gene expression that are modulated through pathways distinct from the drugs intended target. These unique sequences therefore also have utility in defining and monitoring both drug action and toxicity.

As an example of utility, the sequences first disclosed in SEQ ID NOS:1-5 can be utilized in microarrays or other assay formats, to screen collections of genetic material from patients who have a particular medical condition. These investigations can also be carried out using the sequences first disclosed in SEQ ID NOS:1-5 *in silico* and by comparing previously collected genetic databases and the disclosed sequences using computer software known to those in the art.

Thus the sequences first disclosed in SEQ ID NOS:1-5 can be used to identify mutations associated with a particular disease and also as a diagnostic or prognostic assay.

Although the presently described sequences have been specifically described using nucleotide sequence, it should be appreciated that each of the sequences can uniquely be described using any of a wide variety of additional structural attributes, or combinations thereof. For example, a given sequence can be described by the net composition of the nucleotides present within a given region of the sequence in conjunction with the presence of one or more specific oligonucleotide sequence(s) first disclosed in the SEQ ID NOS: 1-5. Alternatively, a restriction map specifying the relative positions of restriction endonuclease digestion sites, or various palindromic or other

specific oligonucleotide sequences can be used to structurally describe a given sequence. Such restriction maps, which are typically generated by widely available computer programs (e.g., the University of Wisconsin GCG sequence analysis package, SEQUENCHER 3.0, Gene Codes Corp., Ann Arbor, MI, etc.), can optionally be used in conjunction with one or more discrete nucleotide sequence(s) present in the sequence that can be described by the relative position of the sequence relative to one or more additional sequence(s) or one or more restriction sites present in the disclosed sequence.

For oligonucleotide probes, highly stringent conditions may refer, e.g., to washing in 6xSSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligos), 48°C (for 17-base oligos), 55°C (for 20-base oligos), and 60°C (for 23-base oligos). These nucleic acid molecules may encode or act as NHP gene antisense molecules, useful, for example, in NHP gene regulation (for and/or as antisense primers in amplification reactions of NHP gene nucleic acid sequences). With respect to NHP gene regulation, such techniques can be used to regulate biological functions. Further, such sequences may be used as part of ribozyme and/or triple helix sequences that are also useful for NHP gene regulation.

Inhibitory antisense or double stranded oligonucleotides can additionally comprise at least one modified base moiety that is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine,

5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 5 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

10 The antisense oligonucleotide can also comprise at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the antisense oligonucleotide 15 will comprise at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

20 In yet another embodiment, the antisense oligonucleotide is an α -anomeric oligonucleotide. An α -anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gautier *et al.*, 1987, Nucl. Acids Res. 15:6625-6641). 25 The oligonucleotide is a 2'-O-methylribonucleotide (Inoue *et al.*, 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue *et al.*, 1987, FEBS Lett. 215:327-330). Alternatively, double stranded RNA can be used to disrupt the expression and function of a targeted NHP.

30 Oligonucleotides of the invention can be synthesized by standard methods known in the art, e.g., by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples,

phosphorothioate oligonucleotides can be synthesized by the method of Stein *et al.* (1988, Nucl. Acids Res. 16:3209), and methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin *et al.*, 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), *etc.*

Low stringency conditions are well known to those of skill in the art, and will vary predictably depending on the specific organisms from which the library and the labeled sequences are derived. For guidance regarding such conditions see, for example, Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual (and periodic updates thereof), Cold Springs Harbor Press, N.Y.; and Ausubel *et al.*, 1989, *supra*.

Alternatively, suitably labeled NHP nucleotide probes can be used to screen a human genomic library using appropriately stringent conditions or by PCR. The identification and characterization of human genomic clones is helpful for identifying polymorphisms (including, but not limited to, nucleotide repeats, microsatellite alleles, single nucleotide polymorphisms, or coding single nucleotide polymorphisms), determining the genomic structure of a given locus/allele, and designing diagnostic tests. For example, sequences derived from regions adjacent to the intron/exon boundaries of the human gene can be used to design primers for use in amplification assays to detect mutations within the exons, introns, splice sites (*e.g.*, splice acceptor and/or donor sites), *etc.*, that can be used in diagnostics and pharmacogenomics.

For example, the present sequences can be used in restriction fragment length polymorphism (RFLP) analysis to identify specific individuals. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification (as generally described in U.S. Pat. No.

5,272,057, incorporated herein by reference). In addition, the sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e., another DNA sequence that is unique to a particular individual). Actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments.

Further, a NHP gene homolog can be isolated from nucleic acid from an organism of interest by performing PCR using two degenerate or "wobble" oligonucleotide primer pools designed on the basis of amino acid sequences within the NHP products disclosed herein. The template for the reaction may be total RNA, mRNA, and/or cDNA obtained by reverse transcription of mRNA prepared from human or non-human cell lines or tissue known or suspected to express an allele of a NHP gene. The PCR product can be subcloned and sequenced to ensure that the amplified sequences represent the sequence of the desired NHP gene. The PCR fragment can then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment can be labeled and used to screen a cDNA library, such as a bacteriophage cDNA library. Alternatively, the labeled fragment can be used to isolate genomic clones via the screening of a genomic library.

PCR technology can also be used to isolate full length cDNA sequences. For example, RNA can be isolated, following standard procedures, from an appropriate cellular or tissue source (i.e., one known, or suspected, to express a NHP gene). A reverse transcription (RT) reaction can be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis.

The resulting RNA/DNA hybrid may then be "tailed" using a standard terminal transferase reaction, the hybrid may be digested with RNase H, and second strand synthesis may then be primed with a complementary primer. Thus, cDNA sequences
5 upstream of the amplified fragment can be isolated. For a review of cloning strategies that can be used, see e.g., Sambrook et al., 1989, *supra*.

A cDNA encoding a mutant NHP sequence can be isolated, for example, by using PCR. In this case, the first cDNA strand may
10 be synthesized by hybridizing an oligo-dT oligonucleotide to mRNA isolated from tissue known or suspected to be expressed in an individual putatively carrying a mutant NHP allele, and by extending the new strand with reverse transcriptase. The second strand of the cDNA is then synthesized using an oligonucleotide
15 that hybridizes specifically to the 5' end of the normal sequence. Using these two primers, the product is then amplified via PCR, optionally cloned into a suitable vector, and subjected to DNA sequence analysis through methods well known to those of skill in the art. By comparing the DNA sequence of the mutant
20 NHP allele to that of a corresponding normal NHP allele, the mutation(s) responsible for the loss or alteration of function of the mutant NHP gene product can be ascertained.

Alternatively, a genomic library can be constructed using DNA obtained from an individual suspected of or known to carry a
25 mutant NHP allele (e.g., a person manifesting a NHP-associated phenotype such as, for example, obesity, high blood pressure, connective tissue disorders, infertility, etc.), or a cDNA library can be constructed using RNA from a tissue known, or suspected, to express a mutant NHP allele. A normal NHP gene, or
30 any suitable fragment thereof, can then be labeled and used as a probe to identify the corresponding mutant NHP allele in such libraries. Clones containing mutant NHP sequences can then be

purified and subjected to sequence analysis according to methods well known to those skilled in the art.

Additionally, an expression library can be constructed utilizing cDNA synthesized from, for example, RNA isolated from a tissue known, or suspected, to express a mutant NHP allele in an individual suspected of or known to carry such a mutant allele. In this manner, gene products made by the putatively mutant tissue can be expressed and screened using standard antibody screening techniques in conjunction with antibodies raised against a normal NHP product, as described below. (For screening techniques, see, for example, Harlow, E. and Lane, eds., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor.) Additionally, screening can be accomplished by screening with labeled NHP fusion proteins, such as, for example, alkaline phosphatase-NHP or NHP-alkaline phosphatase fusion proteins. In cases where a NHP mutation results in an expression product with altered function (e.g., as a result of a missense or a frameshift mutation), polyclonal antibodies to NHP are likely to cross-react with a corresponding mutant NHP expression product. Library clones detected via their reaction with such labeled antibodies can be purified and subjected to sequence analysis according to methods well known in the art.

The invention also encompasses (a) DNA vectors that contain any of the foregoing NHP coding sequences and/or their complements (*i.e.*, antisense); (b) DNA expression vectors that contain any of the foregoing NHP coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences (for example, baculovirus as described in U.S. Patent No. 5,869,336 herein incorporated by reference); (c) genetically engineered host cells that contain any of the foregoing NHP coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences in the host cell; and (d) genetically engineered host

cells that express an endogenous NHP sequence under the control of an exogenously introduced regulatory element (*i.e.*, gene activation). As used herein, regulatory elements include, but are not limited to, inducible and non-inducible promoters, enhancers, operators and other elements known to those skilled in the art that drive and regulate expression. Such regulatory elements include but are not limited to the cytomegalovirus (hCMV) immediate early gene, regulatable, viral elements (particularly retroviral LTR promoters), the early or late promoters of SV40 adenovirus, the *lac* system, the *trp* system, the TAC system, the TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase (PGK), the promoters of acid phosphatase, and the promoters of the yeast α -mating factors.

The present invention also encompasses antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists and agonists of a NHP, as well as compounds or nucleotide constructs that inhibit expression of a NHP sequence (transcription factor inhibitors, antisense and ribozyme molecules, or open reading frame sequence or regulatory sequence replacement constructs), or promote the expression of a NHP (*e.g.*, expression constructs in which NHP coding sequences are operatively associated with expression control elements such as promoters, promoter/enhancers, *etc.*).

The NHPs or NHP peptides, NHP fusion proteins, NHP nucleotide sequences, antibodies, antagonists and agonists can be useful for the detection of mutant NHPs or inappropriately expressed NHPs for the diagnosis of disease. The NHP proteins or peptides, NHP fusion proteins, NHP nucleotide sequences, host cell expression systems, antibodies, antagonists, agonists and genetically engineered cells and animals can be used for screening for drugs (or high throughput screening of

combinatorial libraries) effective in the treatment of the symptomatic or phenotypic manifestations of perturbing the normal function of NHP in the body. The use of engineered host cells and/or animals may offer an advantage in that such systems allow
5 not only for the identification of compounds that bind to the endogenous receptor for an NHP, but can also identify compounds that trigger NHP-mediated activities or pathways.

Finally, the NHP products can be used as therapeutics. For example, soluble derivatives such as NHP peptides/domains
10 corresponding to NHPs, NHP fusion protein products (especially NHP-Ig fusion proteins, *i.e.*, fusions of a NHP, or a domain of a NHP, to an IgFc), NHP antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists or agonists (including compounds that modulate or act on downstream targets in a NHP-
15 mediated pathway) can be used to directly treat diseases or disorders. For instance, the administration of an effective amount of soluble NHP, or a NHP-IgFc fusion protein or an anti-idiotypic antibody (or its Fab) that mimics the NHP could activate or effectively antagonize the endogenous NHP receptor.
20 Nucleotide constructs encoding such NHP products can be used to genetically engineer host cells to express such products *in vivo*; these genetically engineered cells function as "bioreactors" in the body delivering a continuous supply of a NHP, a NHP peptide, or a NHP fusion protein to the body. Nucleotide constructs
25 encoding functional NHPs, mutant NHPs, as well as antisense and ribozyme molecules can also be used in "gene therapy" approaches for the modulation of NHP expression. Thus, the invention also encompasses pharmaceutical formulations and methods for treating biological disorders.

30 Various aspects of the invention are described in greater detail in the subsections below.

5.1 THE NHP SEQUENCES

The cDNA sequences and the corresponding deduced amino acid sequences of the described NHPs are presented in the Sequence Listing. The NHP nucleotides were obtained from clustered human
5 ESTs, and cDNAs from brain and kidney libraries (Edge Biosystems, Gaithersburg, MD). The described NHPs are similar to mammalian ABC transporters and transporters that have been linked to multidrug resistance. Accordingly, the described NHPs can be
10 useful in detecting and treating mental disorders, or in the treatment of cancer.

Several polymorphism were identified during the sequencing of the NHPs, as evidenced by the W (an A/T polymorphism) at position 810 of SEQ ID NOS: 1 and 3 (which results in a tyr or STOP being present at the corresponding amino acid (aa) position
15 270 of SEQ ID NOS:2 and 4); the K (a G/T polymorphism) at position 2494 of SEQ ID NOS: 1 and 3 (which results in a ser or ala being present at the corresponding aa position 832 of SEQ ID NOS:2 and 4); the R (a G/A polymorphism) at position 2878 of SEQ ID NOS: 1 and 3 (which results in a val or met being present at
20 the corresponding aa position 960 of SEQ ID NOS:2 and 4; and a C/T polymorphism at position 3265 of SEQ ID NOS: 1 and 3 (which does not result in a change in the amino acid sequence of SEQ ID NOS:2 and 4, as leu is encoded in both cases).

ABC transporters and transporter related multidrug
25 resistance (MDR) sequences, as well as uses and applications that are germane to the described NHPs, are described in U.S. Patents Nos. 5,198,344, 5,866,699, and 6,080,842, which are herein incorporated by reference in their entirety.

An additional application of the described novel human
30 polynucleotide sequences is their use in the molecular mutagenesis/evolution of proteins that are at least partially encoded by the described novel sequences using, for example, polynucleotide shuffling or related methodologies. Such

approaches are described in U.S. Patents Nos. 5,830,721 and 5,837,458, which are herein incorporated by reference in their entirety.

NHP gene products can also be expressed in transgenic
5 animals. Animals of any species, including, but not limited to, worms, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, birds, goats, and non-human primates, e.g., baboons, monkeys, and chimpanzees may be used to generate NHP transgenic animals.

Any technique known in the art may be used to introduce a
10 NHP transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe, P.C. and Wagner, T.E., 1989, U.S. Pat. No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten et al., 1985, Proc. Natl. Acad. Sci.,
15 USA 82:6148-6152); gene targeting in embryonic stem cells (Thompson et al., 1989, Cell 56:313-321); electroporation of embryos (Lo, 1983, Mol Cell. Biol. 3:1803-1814); and sperm-mediated gene transfer (Lavitrano et al., 1989, Cell 57:717-723); etc. For a review of such techniques, see Gordon, 1989,
20 Transgenic Animals, Intl. Rev. Cytol. 115:171-229, which is incorporated by reference herein in its entirety.

The present invention provides for transgenic animals that carry the NHP transgene in all their cells, as well as animals that carry the transgene in some, but not all their cells, i.e.,
25 mosaic animals or somatic cell transgenic animals. The transgene may be integrated as a single transgene or in concatamers, e.g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et
30 al., 1992, Proc. Natl. Acad. Sci. USA 89:6232-6236. The regulatory sequences required for such a cell-type specific

activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

When it is desired that a NHP transgene be integrated into the chromosomal site of the endogenous NHP gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous NHP gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous NHP gene (*i.e.*, "knockout" animals).

The transgene can also be selectively introduced into a particular cell type, thus inactivating the endogenous NHP gene in only that cell type, by following, for example, the teaching of Gu *et al.*, 1994, *Science*, 265:103-106. The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant NHP gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques that include but are not limited to Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and RT-PCR. Samples of NHP gene-expressing tissue, may also be evaluated immunocytochemically using antibodies specific for the NHP transgene product.

30

5.2 NHPS AND NHP POLYPEPTIDES

NHPs, NHP polypeptides, NHP peptide fragments, mutated, truncated, or deleted forms of the NHPs, and/or NHP fusion

proteins can be prepared for a variety of uses. These uses include, but are not limited to, the generation of antibodies, as reagents in diagnostic assays, for the identification of other cellular gene products related to a NHP, as reagents in assays
5 for screening for compounds that can be used as pharmaceutical reagents useful in the therapeutic treatment of mental, biological, or medical disorders and disease. Given the similarity information and expression data, the described NHPs can be targeted (by drugs, oligos, antibodies, etc.,) in order to
10 treat disease, or to therapeutically augment the efficacy of, for example, chemotherapeutic agents used in the treatment of cancer.

The Sequence Listing discloses the amino acid sequences encoded by the described NHP genes. The NHPs typically display have initiator methionines in DNA sequence contexts consistent
15 with a translation initiation site, and a signal like sequence near the N-terminal regions of the proteins.

The NHP amino acid sequences of the invention include the amino acid sequence presented in the Sequence Listing as well as analogues and derivatives thereof. Further, corresponding NHP
20 homologues from other species are encompassed by the invention. In fact, any NHP protein encoded by the NHP nucleotide sequences described above are within the scope of the invention, as are any novel polynucleotide sequences encoding all or any novel portion of an amino acid sequence presented in the Sequence Listing. The
25 degenerate nature of the genetic code is well known, and, accordingly, each amino acid presented in the Sequence Listing, is generically representative of the well known nucleic acid "triplet" codon, or in many cases codons, that can encode the amino acid. As such, as contemplated herein, the amino acid
30 sequences presented in the Sequence Listing, when taken together with the genetic code (see, for example, Table 4-1 at page 109 of "Molecular Cell Biology", 1986, J. Darnell et al. eds., Scientific American Books, New York, NY, herein incorporated by

reference) are generically representative of all the various permutations and combinations of nucleic acid sequences that can encode such amino acid sequences.

The invention also encompasses proteins that are

5 functionally equivalent to the NHPs encoded by the presently described nucleotide sequences as judged by any of a number of criteria, including, but not limited to, the ability to bind and cleave a substrate of a NHP, or the ability to effect an identical or complementary downstream pathway, or a change in

10 cellular metabolism (e.g., proteolytic activity, ion flux, tyrosine phosphorylation, etc.). Such functionally equivalent NHP proteins include, but are not limited to, additions or substitutions of amino acid residues within the amino acid sequence encoded by the NHP nucleotide sequences described above,

15 but that result in a silent change, thus producing a functionally equivalent expression product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino

20 acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged

25 (acidic) amino acids include aspartic acid and glutamic acid.

A variety of host-expression vector systems can be used to express the NHP nucleotide sequences of the invention. Where, as in the present instance, the NHP peptide or polypeptide is thought to be membrane protein, the hydrophobic regions of the

30 protein can be excised and the resulting soluble peptide or polypeptide can be recovered from the culture media. Such expression systems also encompass engineered host cells that express a NHP, or functional equivalent, *in situ*. Purification

or enrichment of a NHP from such expression systems can be accomplished using appropriate detergents and lipid micelles and methods well known to those skilled in the art. However, such engineered host cells themselves may be used in situations where
5 it is important not only to retain the structural and functional characteristics of the NHP, but to assess biological activity, e.g., in drug screening assays.

The expression systems that may be used for purposes of the invention include but are not limited to microorganisms such as
10 bacteria (e.g., *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing NHP nucleotide sequences; yeast (e.g., *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing NHP nucleotide sequences; insect
15 cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing NHP sequences; plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti
20 plasmid) containing NHP nucleotide sequences; or mammalian cell systems (e.g., COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the
25 vaccinia virus 7.5K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the NHP product being expressed. For example, when a large quantity of such a protein is to be produced for the generation of
30 pharmaceutical compositions of or containing NHP, or for raising antibodies to a NHP, vectors that direct the expression of high levels of fusion protein products that are readily purified may

be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther *et al.*, 1983, EMBO J. 2:1791), in which a NHP coding sequence may be ligated individually into the vector in frame with the *lacZ* coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, 1985, Nucleic Acids Res. 13:3101-3109; Van Heeke & Schuster, 1989, J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors (Pharmacia or American Type Culture Collection) can also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target expression product can be released from the GST moiety.

In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign polynucleotide sequences. The virus grows in *Spodoptera frugiperda* cells. A NHP coding sequence can be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of NHP coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted sequence is expressed (*e.g.*, see Smith *et al.*, 1983, J. Virol. 46: 584; Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as

an expression vector, the NHP nucleotide sequence of interest may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric sequence may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing a NHP product in infected hosts (e.g., See Logan & Shenk, 1984, Proc. Natl. Acad. Sci. USA 81:3655-3659). Specific initiation signals may also be required for efficient translation of inserted NHP nucleotide sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire NHP gene or cDNA, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of a NHP coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (See Bitter et al., 1987, Methods in Enzymol. 153:516-544).

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the expression product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Different host cells have

characteristic and specific mechanisms for the post-translational processing and modification of proteins and expression products. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein
5 expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the expression product may be used. Such mammalian host cells include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38,
10 and in particular, human cell lines.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express the NHP sequences described above can be engineered. Rather than using expression vectors that contain
15 viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (e.g., promoter, enhancer sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be
20 allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci, which in turn can be cloned
25 and expanded into cell lines. This method may advantageously be used to engineer cell lines that express the NHP product. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the NHP product.

30 A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, et al., 1977, Cell 11:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska & Szybalski, 1962, Proc.

Natl. Acad. Sci. USA 48:2026), and adenine phosphoribosyltransferase (Lowy, et al., 1980, Cell 22:817) genes, which can be employed in tk⁻, hgp⁺ or ap⁺ cells, respectively. Also, antimetabolite resistance can be used as the
5 basis of selection for the following genes: dhfr, which confers resistance to methotrexate (Wigler, et al., 1980, Natl. Acad. Sci. USA 77:3567; O'Hare, et al., 1981, Proc. Natl. Acad. Sci. USA 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan & Berg, 1981, Proc. Natl. Acad. Sci. USA 78:2072); neo,
10 which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, et al., 1981, J. Mol. Biol. 150:1); and hyg⁺, which confers resistance to hygromycin (Santerre, et al., 1984, Gene 30:147).

Alternatively, any fusion protein can be readily purified
15 by utilizing an antibody specific for the fusion protein being expressed. For example, a system described by Janknecht et al. allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht, et al., 1991, Proc. Natl. Acad. Sci. USA 88:8972-8976). In this system, the
20 sequence of interest is subcloned into a vaccinia recombination plasmid such that the sequence's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni²⁺-nitriloacetic
25 acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

Also encompassed by the present invention are fusion proteins that direct the NHP to a target organ and/or facilitate transport across the membrane into the cytosol. Conjugation of
30 NHPs to antibody molecules or their Fab fragments could be used to target cells bearing a particular epitope. Attaching the appropriate signal sequence to the NHP would also transport the

NHP to the desired location within the cell. Alternatively targeting of NHP or its nucleic acid sequence might be achieved using liposome or lipid complex based delivery systems. Such technologies are described in "Liposomes: A Practical Approach",
5 New, R.R.C., ed., Oxford University Press, New York and in U.S. Patents Nos. 4,594,595, 5,459,127, 5,948,767 and 6,110,490 and their respective disclosures, which are herein incorporated by reference in their entirety. Additionally embodied are novel protein constructs engineered in such a way that they facilitate
10 transport of the NHP to the target site or desired organ, where they cross the cell membrane and/or the nucleus where the NHP can exert its functional activity. This goal may be achieved by coupling of the NHP to a cytokine or other ligand that provides targeting specificity, and/or to a protein transducing domain
15 (see generally U.S. applications Ser. No. 60/111,701 and 60/056,713, both of which are herein incorporated by reference, for examples of such transducing sequences) to facilitate passage across cellular membranes and can optionally be engineered to include nuclear localization.

20

5.3 ANTIBODIES TO NHP PRODUCTS

Antibodies that specifically recognize one or more epitopes of a NHP, or epitopes of conserved variants of a NHP, or peptide fragments of a NHP are also encompassed by the invention. Such
25 antibodies include but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')₂ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the
30 above.

The antibodies of the invention may be used, for example, in the detection of NHP in a biological sample and may, therefore, be utilized as part of a diagnostic or prognostic

technique whereby patients may be tested for abnormal amounts of NHP. Such antibodies may also be utilized in conjunction with, for example, compound screening schemes for the evaluation of the effect of test compounds on expression and/or activity of a NHP expression product. Additionally, such antibodies can be used in conjunction gene therapy to, for example, evaluate the normal and/or engineered NHP-expressing cells prior to their introduction into the patient. Such antibodies may additionally be used as a method for the inhibition of abnormal NHP activity. Thus, such antibodies may, therefore, be utilized as part of treatment methods.

For the production of antibodies, various host animals may be immunized by injection with a NHP, an NHP peptide (e.g., one corresponding to a functional domain of an NHP), truncated NHP polypeptides (NHP in which one or more domains have been deleted), functional equivalents of the NHP or mutated variant of the NHP. Such host animals may include but are not limited to pigs, rabbits, mice, goats, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's adjuvant (complete and incomplete), mineral salts such as aluminum hydroxide or aluminum phosphate, chitosan, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Alternatively, the immune response could be enhanced by combination and or coupling with molecules such as keyhole limpet hemocyanin, tetanus toxoid, diphtheria toxoid, ovalbumin, cholera toxin or fragments thereof. Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of the immunized animals.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, can be obtained by any

technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler and Milstein, (1975, *Nature* 256:495-497; and U.S. Patent No. 4,376,110), the
5 human B-cell hybridoma technique (Kosbor et al., 1983, *Immunology Today* 4:72; Cole et al., 1983, *Proc. Natl. Acad. Sci. USA* 80:2026-2030), and the EBV-hybridoma technique (Cole et al., 1985, *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin
10 class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated *in vitro* or *in vivo*. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production.

15 In addition, techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, *Proc. Natl. Acad. Sci.*, 81:6851-6855; Neuberger et al., 1984, *Nature*, 312:604-608; Takeda et al., 1985, *Nature*, 314:452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity
20 together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region.
25 Such technologies are described in U.S. Patents Nos. 6,075,181 and 5,877,397 and their respective disclosures, which are herein incorporated by reference in their entirety. Also encompassed by the present invention is the use of fully humanized monoclonal antibodies as described in US Patent No. 6,150,584 and respective
30 disclosures, which are herein incorporated by reference in their entirety.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778; Bird, 1988, Science 242:423-426; Huston et al., 1988, Proc. Natl. Acad. Sci. USA 85:5879-5883; and Ward et al., 1989, Nature 341:544-546) can
5 be adapted to produce single chain antibodies against NHP expression products. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments that recognize specific epitopes may be
10 generated by known techniques. For example, such fragments include, but are not limited to: the F(ab')₂ fragments, which can be produced by pepsin digestion of the antibody molecule and the Fab fragments, which can be generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression
15 libraries may be constructed (Huse et al., 1989, Science, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Antibodies to a NHP can, in turn, be utilized to generate anti-idiotypic antibodies that "mimic" a given NHP, using
20 techniques well known to those skilled in the art. (See, e.g., Greenspan & Bona, 1993, FASEB J 7(5):437-444; and Nissinoff, 1991, J. Immunol. 147(8):2429-2438). For example antibodies that bind to a NHP domain and competitively inhibit the binding of NHP to its cognate receptor can be used to generate anti-idiotypes
25 that "mimic" the NHP and, therefore, bind and activate or neutralize a receptor. Such anti-idiotypic antibodies or Fab fragments of such anti-idiotypes can be used in therapeutic regimens involving a NHP mediated pathway.

Additionally given the high degree of relatedness of
30 mammalian NHPs, the presently described knock-out mice (having never seen NHP, and thus never been tolerized to NHP) have a unique utility, as they can be advantageously applied to the

generation of antibodies against the disclosed mammalian NHP (i.e., NHP will be immunogenic in NHP knock-out animals).

The present invention is not to be limited in scope by the specific embodiments described herein, which are intended as
5 single illustrations of individual aspects of the invention, and functionally equivalent methods and components are within the scope of the invention. Indeed, various modifications of the invention, in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing
10 description. Such modifications are intended to fall within the scope of the appended claims. All cited publications, patents, and patent applications are herein incorporated by reference in their entirety.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising a nucleotide sequence drawn from the group consisting of SEQ ID
5 NO:1 and SEQ ID NO:3.
2. An isolated nucleic acid molecule comprising a nucleotide sequence that:
 - (a) encodes the amino acid sequence shown in SEQ ID
10 NO:2; and
 - (b) hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO:1 or the complement thereof.
- 15 3. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes the amino acid sequence shown in SEQ ID NO:2.
- 20 4. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes the amino acid sequence shown in SEQ ID NO:4.

SEQUENCE LISTING

<110> LEXICON GENETICS INCORPORATED

<120> Novel Human Transporter Proteins and Polynucleotides Encoding the Same

<130> LEX-0250-PCT

<150> US 60/239,629

<151> 2000-10-10

<160> 5

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 4929

<212> DNA

<213> homo sapiens

<400> 1

atgtccactg	caattagggg	ggtaggagtt	tggagacaga	ccagaacact	tctactgaag	60
aattacttaa	ttaaatgcag	aaccaaaaaag	agtagtggtc	aggaaattct	ttttccacta	120
ttttttttat	tttgggttaat	attaattagc	atgatgcata	caaataagaa	atatgaagaa	180
gtgcctaata	tagaactcaa	tcctatggac	aagtttactc	tttctaattc	aattcttgga	240
tatactccag	tgactaatat	tacaagcagc	atcatgcaga	aagtgtctac	tgatcatcta	300
cctgatgtca	taattactga	agaatataca	aatgaaaaag	aaatgttaac	atccagtctc	360
tctaagccga	gcaactttgt	aggtgtggtt	ttcaaagact	ccatgtccta	tgaacttcgt	420
ttttttcctg	atatgattcc	agtatcttct	atttatatgg	attcaagagc	tggctgttca	480
aaatcatgtg	aggctgctca	gtactggctc	tcaggtttca	cagttttaca	agcatccata	540
gatgctgcca	ttatacagtt	gaagaccaat	gtttctcttt	ggaaggagct	ggagtcaact	600
aaagctgtta	ttatgggaga	aactgctggt	gtagaaatag	atacctttcc	ccgaggagta	660
attttaatat	acctagttat	agcattttca	ccttttggat	actttttggc	aattcataatc	720
gtagcagaaa	aagaaaaaaa	aataaaagaa	tttttaaaga	taatgggact	tcatagatact	780
gccttttggc	tttctctggg	tcttctataw	acaagtttaa	tttttcttat	gtcccttctt	840
atggcagtc	ttgcgacagc	ttctttgtta	tttctctcaa	gtagcagcat	tgtgatattt	900
ctgctttttt	tccttttatgg	attatcatct	gtattttttg	ctttaatgct	gacacctctt	960
tttaaaaaat	caaaacatgt	gggaatagtt	gaattttttg	ttactgtggc	ttttggattt	1020
attggcctta	tgataatcct	catagaaagt	tttcccaa	cgtagtgtg	gcttttcagt	1080
cctttctgtc	actgtacttt	tgtgattggt	attgcacagg	tcatgcattt	agaagatttt	1140
aatgaagggt	cttcattttt	aaatttgact	gcaggcccat	atcctcta	tattacaatt	1200
atcatgctca	cacttaatat	tatatcttat	gtcctcttgg	ctgtctatct	tgatcaagtc	1260
attccagggg	aatttggtct	acggagatca	tctttatatt	ttctgaagcc	ttcataattg	1320
tcaaagagca	aaagaaatta	tgaggagtta	tcagagggca	atgttaatgg	aaatattagt	1380
tttagtgaaa	ttattgagcc	agtttcttca	gaattttag	gaaaagaagc	cataagaatt	1440
agtggtatcc	agaagacata	cagaaagaag	ggtgaaaatg	tggaggcttt	gagaaatttg	1500
tcatttgaca	tatatgaggg	tcagattact	gccttacttg	gccacagtgg	aacaggaaag	1560
agtacattga	tgaatattct	ttgtggactc	tgcccacctt	ctgatgggtt	tgcatctata	1620
tatggacaca	gagtctcaga	aatagatgaa	atgtttgaag	caagaaaaat	gattggcatt	1680
tgtccacagt	tagatataca	ctttgatggt	ttgacagtag	aagaaaaatt	atcaattttg	1740
gcttcaatca	aagggatacc	agccaccaat	ataatacaag	aagtgcagaa	ggttttacta	1800
gatttagaca	tgcagactat	caaagataac	caagctaaaa	aattaagtgg	tgggtcaaaa	1860
agaaagctgt	cattaggaat	tgctgttctt	gggaacccaa	agatactgct	gctagatgaa	1920
ccaacagctg	gaatggaccc	ctgttctcga	catattgtat	ggaatctttt	aaaatacaga	1980

aaagccaatc	gggtgacagt	gttcagtact	catttcatgg	atgaagctga	cattcttgca	2040
gataggaaag	ctgtgatata	acaaggaatg	ctgaaatgtg	ttggttcttc	aatgttcctc	2100
aaaagtaa	at	ctacgcctg	agcatgtaca	tagacaaata	ttgtgccaca	2160
gaatctcttt	cttcactgg	tacaacacat	atacctggag	ctactttatt	acaacagaat	2220
gaccaacaac	ttgtgtatag	cttgcccttc	aaggacatgg	acaaatcttc	aggtttggtt	2280
tctgccctag	acagtcattc	aaatttgggt	ggcatttctt	atgggggttc	catgacgact	2340
ttggaagacg	tatttttaaa	gctagaagtt	gaagcagaaa	ttgaccaagc	agattatagt	2400
gtatttactc	agcagccact	ggaggaagaa	atggattcaa	aatcttttga	tgaaatggaa	2460
cagagccttac	ttattctttc	tgaaaccaag	gctkctctag	tgagcaccat	gagcctttgg	2520
aaacaacaga	tgtatacaat	agcaaagttt	catttcttta	ccttgaaacg	tgaaagtaaa	2580
tcagtggat	cagtgttgct	tctgctttta	atctttttca	cagttcagat	ttttatgttt	2640
ttgggttcac	actcttttaa	aaatgctgtg	gttcccatca	aacttggtcc	agacttatat	2700
tttctaaaac	ctggagacaa	accacataaa	tacaaaacaa	gtctgcttct	tcaaaattct	2760
gctgactcag	atatcagtga	tcttattagc	tttttcacaa	gccagaacat	aatgggtgacg	2820
atgattaatg	acagtgacta	tgtatccgtg	gtcctccata	gtgcggcttt	aaatgtgrtg	2880
cattcagaaa	aggactatgt	ttttgcagct	gttttcaaca	gtactatggg	ttattcttta	2940
cctatatattg	tgaatatcat	tagtaactac	tatctttatc	atttaaagt	gactgaaacc	3000
atccagatct	ggagtacccc	attctttcaa	gaaattactg	atatagtttt	taaaattgag	3060
ctgtattttc	aagcagcttt	gcttggaa	attgttactg	caatgccacc	ttactttgcc	3120
atggaaaatg	cagagaatca	taagatcaaa	gcttayactc	aacttaaa	ttcagggtctt	3180
ttgccatctg	catattggat	tggaacaagct	gttgttgata	tccccttatt	ttttatcatt	3240
cttattttga	tgctaggaag	cttatttgca	tttcattatg	gattatattt	ttatactgta	3300
aagttccttg	ctgtgggttt	ttgccttatt	ggttatgttc	catcagttat	tctgttcact	3360
tataattgctt	ctttcacctt	taagaaaatt	ttaaatacca	aagaattttg	gtcattttatc	3420
tattctgtgg	cagcgttggc	ttgtattgca	atcactgaaa	taactttctt	tatgggatac	3480
acaattgcaa	ctattcttca	ttatgccttt	tgtatcatca	ttccaatcta	tccacttcta	3540
ggttgcctga	tttctttcat	aaagatttct	tggaagaatg	tacgaaaaaa	tgtggacacc	3600
tataatccat	gggataggct	ttcagtagct	gttatatcgc	cttacctgca	gtgtgtactg	3660
tggaattttcc	tcttacaata	ctatgagaaa	aaatatggag	gcagatcaat	aagaaaagat	3720
ccctttttca	gaaacctttc	aacgaagtct	aaaaatagga	agcttccaga	accaccagac	3780
aatgaggatg	aagatgaaga	tgtcaaaagct	gaaagactaa	aggtaaaga	gctgatgggt	3840
tgccagtggt	gtgaggagaa	accatccatt	atggctagca	atttgcataa	agaatatgat	3900
gacaagaaag	atcttcttct	ttcaagaaaa	gtaaagaaag	tggaactaa	atacatctct	3960
ttctgtgtga	aaaaaggaga	gatcttagga	ctattgggtc	caaatgggtg	tggaacaaag	4020
acaattatta	atattctggt	tggtgatatt	gaaccaactt	caggccagggt	attttttagga	4080
gattattctt	cagagacaag	tgaagatgat	gattcactga	agtgtatggg	ttactgtcct	4140
cagataaacc	ctttgtggcc	agatactaca	ttgcaggaac	attttgaaat	ttatggagct	4200
gtcaaaggaa	tgagtgaag	tgacatgaaa	gaagtcataa	gtcgaataac	acatgcactt	4260
gatttaaaag	aacatcttca	gaagactgta	aagaaactac	ctgcaggaat	caaacgaaag	4320
ttgtgttttg	ctctaagtat	gctagggaa	cctcagatta	ctttgctaga	tgaaccatct	4380
acaggatagg	atcccaaagc	caaacagcac	atgtggcgag	caattcgaac	tgcaatttaa	4440
aacagaaaagc	gggctgctat	tctgaccact	cactatatgg	aggaggcaga	ggctgtctgt	4500
gatcgagtag	ctatcatggt	gtctgggcag	ttaagatgta	tcggaacagt	acaacatcta	4560
aagagtaaat	ttggaaaagg	ctactttttg	gaaattaaat	tgaaggactg	gatagaaaac	4620
ctagaagtag	accgccttca	aagagaaatt	cagtataatt	tcccaaatgc	aagccgtcag	4680
gaaagttttt	cttctatttt	ggcttataaa	attcctaagg	aagatgttca	gtccctttca	4740
caatcttttt	ttaaagctgga	agaagctaaa	catgcttttg	ccattgaaga	atatagcttt	4800
tctcaagcaa	cattggaaca	ggtttttgtg	gaactcacta	aagaacaaga	ggaggaagat	4860
aatagttgtg	gaacttttaa	cagcacactt	tggtgggaac	gaacacaaga	agatagagta	4920
gtattttga						4929

<210> 2

<211> 1642

<212> PRT

<213> homo sapiens

<400> 2

```

Met Ser Thr Ala Ile Arg Glu Val Gly Val Trp Arg Gln Thr Arg Thr
 1           5           10           15
Leu Leu Leu Lys Asn Tyr Leu Ile Lys Cys Arg Thr Lys Lys Ser Ser
      20           25           30
Val Gln Glu Ile Leu Phe Pro Leu Phe Phe Leu Phe Trp Leu Ile Leu
      35           40           45
Ile Ser Met Met His Pro Asn Lys Lys Tyr Glu Glu Val Pro Asn Ile
      50           55           60
Glu Leu Asn Pro Met Asp Lys Phe Thr Leu Ser Asn Leu Ile Leu Gly
      65           70           75           80
Tyr Thr Pro Val Thr Asn Ile Thr Ser Ser Ile Met Gln Lys Val Ser
      85           90           95
Thr Asp His Leu Pro Asp Val Ile Ile Thr Glu Glu Tyr Thr Asn Glu
      100          105          110
Lys Glu Met Leu Thr Ser Ser Leu Ser Lys Pro Ser Asn Phe Val Gly
      115          120          125
Val Val Phe Lys Asp Ser Met Ser Tyr Glu Leu Arg Phe Phe Pro Asp
      130          135          140
Met Ile Pro Val Ser Ser Ile Tyr Met Asp Ser Arg Ala Gly Cys Ser
      145          150          155          160
Lys Ser Cys Glu Ala Ala Gln Tyr Trp Ser Ser Gly Phe Thr Val Leu
      165          170          175
Gln Ala Ser Ile Asp Ala Ala Ile Ile Gln Leu Lys Thr Asn Val Ser
      180          185          190
Leu Trp Lys Glu Leu Glu Ser Thr Lys Ala Val Ile Met Gly Glu Thr
      195          200          205
Ala Val Val Glu Ile Asp Thr Phe Pro Arg Gly Val Ile Leu Ile Tyr
      210          215          220
Leu Val Ile Ala Phe Ser Pro Phe Gly Tyr Phe Leu Ala Ile His Ile
      225          230          235          240
Val Ala Glu Lys Glu Lys Lys Ile Lys Glu Phe Leu Lys Ile Met Gly
      245          250          255
Leu His Asp Thr Ala Phe Trp Leu Ser Trp Val Leu Leu Tyr Thr Ser
      260          265          270
Leu Ile Phe Leu Met Ser Leu Leu Met Ala Val Ile Ala Thr Ala Ser
      275          280          285
Leu Leu Phe Pro Gln Ser Ser Ser Ile Val Ile Phe Leu Leu Phe Phe
      290          295          300
Leu Tyr Gly Leu Ser Ser Val Phe Phe Ala Leu Met Leu Thr Pro Leu
      305          310          315          320
Phe Lys Lys Ser Lys His Val Gly Ile Val Glu Phe Phe Val Thr Val
      325          330          335
Ala Phe Gly Phe Ile Gly Leu Met Ile Ile Leu Ile Glu Ser Phe Pro
      340          345          350
Lys Ser Leu Val Trp Leu Phe Ser Pro Phe Cys His Cys Thr Phe Val
      355          360          365
Ile Gly Ile Ala Gln Val Met His Leu Glu Asp Phe Asn Glu Gly Ala
      370          375          380
Ser Phe Ser Asn Leu Thr Ala Gly Pro Tyr Pro Leu Ile Ile Thr Ile
      385          390          395          400
Ile Met Leu Thr Leu Asn Ser Ile Phe Tyr Val Leu Leu Ala Val Tyr
      405          410          415
Leu Asp Gln Val Ile Pro Gly Glu Phe Gly Leu Arg Arg Ser Ser Leu
      420          425          430
Tyr Phe Leu Lys Pro Ser Tyr Trp Ser Lys Ser Lys Arg Asn Tyr Glu

```

435	440	445
Glu Leu Ser Glu Gly Asn Val Asn Gly Asn Ile Ser Phe Ser Glu Ile		
450	455	460
Ile Glu Pro Val Ser Ser Glu Phe Val Gly Lys Glu Ala Ile Arg Ile		
465	470	475
Ser Gly Ile Gln Lys Thr Tyr Arg Lys Lys Gly Glu Asn Val Glu Ala		480
485	490	495
Leu Arg Asn Leu Ser Phe Asp Ile Tyr Glu Gly Gln Ile Thr Ala Leu		
500	505	510
Leu Gly His Ser Gly Thr Gly Lys Ser Thr Leu Met Asn Ile Leu Cys		
515	520	525
Gly Leu Cys Pro Pro Ser Asp Gly Phe Ala Ser Ile Tyr Gly His Arg		
530	535	540
Val Ser Glu Ile Asp Glu Met Phe Glu Ala Arg Lys Met Ile Gly Ile		
545	550	555
Cys Pro Gln Leu Asp Ile His Phe Asp Val Leu Thr Val Glu Glu Asn		560
565	570	575
Leu Ser Ile Leu Ala Ser Ile Lys Gly Ile Pro Ala Thr Asn Ile Ile		
580	585	590
Gln Glu Val Gln Lys Val Leu Leu Asp Leu Asp Met Gln Thr Ile Lys		
595	600	605
Asp Asn Gln Ala Lys Lys Leu Ser Gly Gly Gln Lys Arg Lys Leu Ser		
610	615	620
Leu Gly Ile Ala Val Leu Gly Asn Pro Lys Ile Leu Leu Leu Asp Glu		640
625	630	635
Pro Thr Ala Gly Met Asp Pro Cys Ser Arg His Ile Val Trp Asn Leu		655
645	650	655
Leu Lys Tyr Arg Lys Ala Asn Arg Val Thr Val Phe Ser Thr His Phe		
660	665	670
Met Asp Glu Ala Asp Ile Leu Ala Asp Arg Lys Ala Val Ile Ser Gln		
675	680	685
Gly Met Leu Lys Cys Val Gly Ser Ser Met Phe Leu Lys Ser Lys Trp		
690	695	700
Gly Ile Gly Tyr Arg Leu Ser Met Tyr Ile Asp Lys Tyr Cys Ala Thr		720
705	710	715
Glu Ser Leu Ser Ser Leu Val Lys Gln His Ile Pro Gly Ala Thr Leu		
725	730	735
Leu Gln Gln Asn Asp Gln Gln Leu Val Tyr Ser Leu Pro Phe Lys Asp		
740	745	750
Met Asp Lys Phe Ser Gly Leu Phe Ser Ala Leu Asp Ser His Ser Asn		
755	760	765
Leu Gly Gly Ile Ser Tyr Gly Val Ser Met Thr Thr Leu Glu Asp Val		
770	775	780
Phe Leu Lys Leu Glu Val Glu Ala Glu Ile Asp Gln Ala Asp Tyr Ser		800
785	790	795
Val Phe Thr Gln Gln Pro Leu Glu Glu Glu Met Asp Ser Lys Ser Phe		815
805	810	815
Asp Glu Met Glu Gln Ser Leu Leu Ile Leu Ser Glu Thr Lys Ala Ser		
820	825	830
Leu Val Ser Thr Met Ser Leu Trp Lys Gln Gln Met Tyr Thr Ile Ala		
835	840	845
Lys Phe His Phe Phe Thr Leu Lys Arg Glu Ser Lys Ser Val Arg Ser		
850	855	860
Val Leu Leu Leu Leu Leu Ile Phe Phe Thr Val Gln Ile Phe Met Phe		880
865	870	875
Leu Val His His Ser Phe Lys Asn Ala Val Val Pro Ile Lys Leu Val		

885							890							895						
Pro	Asp	Leu	Tyr	Phe	Leu	Lys	Pro	Gly	Asp	Lys	Pro	His	Lys	Tyr	Lys					
900							905							910						
Thr	Ser	Leu	Leu	Leu	Gln	Asn	Ser	Ala	Asp	Ser	Asp	Ile	Ser	Asp	Leu					
915							920							925						
Ile	Ser	Phe	Phe	Thr	Ser	Gln	Asn	Ile	Met	Val	Thr	Met	Ile	Asn	Asp					
930							935							940						
Ser	Asp	Tyr	Val	Ser	Val	Ala	Pro	His	Ser	Ala	Ala	Leu	Asn	Val	Val					
945							950							955						
His	Ser	Glu	Lys	Asp	Tyr	Val	Phe	Ala	Ala	Val	Phe	Asn	Ser	Thr	Met					
965							970							975						
Val	Tyr	Ser	Leu	Pro	Ile	Leu	Val	Asn	Ile	Ile	Ser	Asn	Tyr	Tyr	Leu					
980							985							990						
Tyr	His	Leu	Asn	Val	Thr	Glu	Thr	Ile	Gln	Ile	Trp	Ser	Thr	Pro	Phe					
995							1000							1005						
Phe	Gln	Glu	Ile	Thr	Asp	Ile	Val	Phe	Lys	Ile	Glu	Leu	Tyr	Phe	Gln					
1010							1015							1020						
Ala	Ala	Leu	Leu	Gly	Ile	Ile	Val	Thr	Ala	Met	Pro	Pro	Tyr	Phe	Ala					
1025							1030							1035						
Met	Glu	Asn	Ala	Glu	Asn	His	Lys	Ile	Lys	Ala	Tyr	Thr	Gln	Leu	Lys					
1045							1050							1055						
Leu	Ser	Gly	Leu	Leu	Pro	Ser	Ala	Tyr	Trp	Ile	Gly	Gln	Ala	Val	Val					
1060							1065							1070						
Asp	Ile	Pro	Leu	Phe	Phe	Ile	Ile	Leu	Ile	Leu	Met	Leu	Gly	Ser	Leu					
1075							1080							1085						
Leu	Ala	Phe	His	Tyr	Gly	Leu	Tyr	Phe	Tyr	Thr	Val	Lys	Phe	Leu	Ala					
1090							1095							1100						
Val	Val	Phe	Cys	Leu	Ile	Gly	Tyr	Val	Pro	Ser	Val	Ile	Leu	Phe	Thr					
1105							1110							1115						
Tyr	Ile	Ala	Ser	Phe	Thr	Phe	Lys	Lys	Ile	Leu	Asn	Thr	Lys	Glu	Phe					
1125							1130							1135						
Trp	Ser	Phe	Ile	Tyr	Ser	Val	Ala	Ala	Leu	Ala	Cys	Ile	Ala	Ile	Thr					
1140							1145							1150						
Glu	Ile	Thr	Phe	Phe	Met	Gly	Tyr	Thr	Ile	Ala	Thr	Ile	Leu	His	Tyr					
1155							1160							1165						
Ala	Phe	Cys	Ile	Ile	Ile	Pro	Ile	Tyr	Pro	Leu	Leu	Gly	Cys	Leu	Ile					
1170							1175							1180						
Ser	Phe	Ile	Lys	Ile	Ser	Trp	Lys	Asn	Val	Arg	Lys	Asn	Val	Asp	Thr					
1185							1190							1195						
Tyr	Asn	Pro	Trp	Asp	Arg	Leu	Ser	Val	Ala	Val	Ile	Ser	Pro	Tyr	Leu					
1205							1210							1215						
Gln	Cys	Val	Leu	Trp	Ile	Phe	Leu	Leu	Gln	Tyr	Tyr	Glu	Lys	Lys	Tyr					
1220							1225							1230						
Gly	Gly	Arg	Ser	Ile	Arg	Lys	Asp	Pro	Phe	Phe	Arg	Asn	Leu	Ser	Thr					
1235							1240							1245						
Lys	Ser	Lys	Asn	Arg	Lys	Leu	Pro	Glu	Pro	Pro	Asp	Asn	Glu	Asp	Glu					
1250							1255							1260						
Asp	Glu	Asp	Val	Lys	Ala	Glu	Arg	Leu	Lys	Val	Lys	Glu	Leu	Met	Gly					
1265							1270							1275						
Cys	Gln	Cys	Cys	Glu	Glu	Lys	Pro	Ser	Ile	Met	Val	Ser	Asn	Leu	His					
1285							1290							1295						
Lys	Glu	Tyr	Asp	Asp	Lys	Lys	Asp	Phe	Leu	Leu	Ser	Arg	Lys	Val	Lys					
1300							1305							1310						
Lys	Val	Ala	Thr																	

1330	1335	1340
Ile Leu Val Gly Asp	Ile Glu Pro Thr Ser Gly Gln Val Phe Leu Gly	
1345	1350	1355
Asp Tyr Ser Ser Glu Thr Ser Glu Asp Asp Ser Leu Lys Cys Met		1360
	1365	1370
Gly Tyr Cys Pro Gln Ile Asn Pro Leu Trp Pro Asp Thr Thr Leu Gln		1375
	1380	1385
Glu His Phe Glu Ile Tyr Gly Ala Val Lys Gly Met Ser Ala Ser Asp		1390
	1395	1400
Met Lys Glu Val Ile Ser Arg Ile Thr His Ala Leu Asp Leu Lys Glu		1405
	1410	1415
His Leu Gln Lys Thr Val Lys Lys Leu Pro Ala Gly Ile Lys Arg Lys		1420
1425	1430	1435
Leu Cys Phe Ala Leu Ser Met Leu Gly Asn Pro Gln Ile Thr Leu Leu		1440
	1445	1450
Asp Glu Pro Ser Thr Gly Met Asp Pro Lys Ala Lys Gln His Met Trp		1455
	1460	1465
Arg Ala Ile Arg Thr Ala Phe Lys Asn Arg Lys Arg Ala Ala Ile Leu		1470
	1475	1480
Thr Thr His Tyr Met Glu Glu Ala Glu Ala Val Cys Asp Arg Val Ala		1485
	1490	1495
Ile Met Val Ser Gly Gln Leu Arg Cys Ile Gly Thr Val Gln His Leu		1500
1505	1510	1515
Lys Ser Lys Phe Gly Lys Gly Tyr Phe Leu Glu Ile Lys Leu Lys Asp		1520
	1525	1530
Trp Ile Glu Asn Leu Glu Val Asp Arg Leu Gln Arg Glu Ile Gln Tyr		1535
	1540	1545
Ile Phe Pro Asn Ala Ser Arg Gln Glu Ser Phe Ser Ser Ile Leu Ala		1550
	1555	1560
Tyr Lys Ile Pro Lys Glu Asp Val Gln Ser Leu Ser Gln Ser Phe Phe		1565
	1570	1575
Lys Leu Glu Glu Ala Lys His Ala Phe Ala Ile Glu Glu Tyr Ser Phe		1580
1585	1590	1595
Ser Gln Ala Thr Leu Glu Gln Val Phe Val Glu Leu Thr Lys Glu Gln		1600
	1605	1610
Glu Glu Glu Asp Asn Ser Cys Gly Thr Leu Asn Ser Thr Leu Trp Trp		1615
	1620	1625
Glu Arg Thr Gln Glu Asp Arg Val Val Phe		1630
	1635	1640

<210> 3

<211> 4785

<212> DNA

<213> homo sapiens

<400> 3

atgtccactg	caattagga	ggtaggagtt	tggagacaga	ccagaacact	tctactgaag	60
aattacttaa	ttaaatagcag	aacaaaaag	agtagtggtc	aggaaattct	ttttccacta	120
ttttttttat	tttggttaat	attaattagc	atgatgcac	caaataagaa	atatgaagaa	180
gtgcctaata	tagaactcaa	tcctatggac	aagtttactc	tttctaactc	aattcttgga	240
tatactccag	tgactaatat	tacaagcagc	atcatgcaga	aagtgtctac	tgatcatcta	300
cctgatgtca	taattactga	agaatataca	aataaaaaag	aaatgttaac	atccagtctc	360
tctaagccga	gcaactttgt	aggtgtggtt	ttcaaagact	ccatgtccta	tgaacttcgt	420
ttttttcctg	atatgattcc	agtatcttct	atttatatgg	attcaagagc	tggctgttca	480
aaatcatgtg	aggtgtctca	gtactggtcc	tcagggttca	cagttttaca	agcatccata	540
gatgctgcca	ttatacagtt	gaagaccaat	gtttctcttt	ggaaggagct	ggagtcaact	600

aaagctgtta	ttatgggaga	aactgctgtt	gtagaaatag	atacctttcc	ccgaggagta	660
attttaatat	acctagttaa	agcattttca	ccttttggat	actttttggc	aattcatatc	720
gtagcagaaa	aagaaaaaaa	aataaaagaa	tttttaaaga	taatgggact	tcatgatact	780
gccttttggc	tttcctgggt	tcttctataw	acaagtttaa	ttttcttat	gtcccttctt	840
atggcagtc	ttgcgacagc	ttctttgtta	tttcctcaaa	gtagcagcat	tgtgatattt	900
ctgctttttt	tccttttatg	attatcatct	gtattttttg	ctttaatgct	gacacctctt	960
tttaaaaaat	caaaacatgt	gggaatagtt	gaattttttg	ttactgtggc	ttttggattt	1020
attggcctta	tgataatcct	catagaaagt	tttcccaaat	cgtagtggtg	gcttttcagt	1080
cctttctgtc	actgtacttt	tgtgattggg	attgcacagg	tcatgcattt	agaagatttt	1140
aatgaagggt	cttcattttc	aaatttgact	gcaggcccat	atcctcta	tattacaatt	1200
atcatgctca	cacttaatat	tatatcttat	gtcctcttgg	ctgtctatct	tgatcaagtc	1260
attccagggg	aatttggcct	acggagatca	tctttatatt	ttctgaagcc	ttcatatttg	1320
tcaaagagca	aaagaaatta	tgaggagtta	tcagagggca	atgttaatgg	aaatattagt	1380
tttagtgaaa	ttattgagcc	agtttcttca	gaatttgtag	gaaaagaagc	cataagaatt	1440
agtggtattc	agaagacata	cagaaaagag	ggtgaaaatg	tggaggcctt	gagaaatttg	1500
tcatttgaca	tatatgaggg	tcagattact	gccttacttg	gccacagtgg	aacaggaaaag	1560
agtacattga	tgaatattct	ttgtggactc	tgcccacctt	ctgatgggtt	tgcatctata	1620
tatggacaca	gagtcctaga	aatagatgaa	atgtttgaag	caagaaaaat	gattggcatt	1680
tgctccacagt	tagatataca	ctttgatgtt	ttgacagtag	aagaaaattt	atcaattttg	1740
gcttcaatca	aagggatacc	agccaccaat	ataatacaag	aagtgcagaa	ggttttacta	1800
gatttagaca	tgacagactat	caaagataac	caagctaaaa	aattaagtgg	tggtcaaaaa	1860
agaaagctgt	catttaggaat	tgctgttctt	gggaacccaa	agatactgct	gctagatgaa	1920
ccaacagctg	gaatggaccc	ctgttctcga	catattgtat	ggaatctttt	aaaatacaga	1980
aaagccaatc	gggtgacagt	gttcagtact	catttcatgg	atgaagctga	cattcttgca	2040
gataggaaa	ctgtgatatc	acaaggaaatg	ctgaaatgtg	ttggttcttc	aatgttcctc	2100
aaaagtaaat	gggggatcgg	ctaccgcctg	agcatgtaca	tagacaaata	ttgtgccaca	2160
gaatctcttt	cttcactggt	taaacacaat	atacctggag	ctactttatt	acaacagaat	2220
gaccaacaac	ttgtgtatag	cttgccttcc	aaggacatgg	acaaattttc	aggtttgttt	2280
tctgccctag	acagtcattc	aaatttgggt	ggcatttctt	atggggtttc	catgacgact	2340
ttggaagacg	tatttttaaa	gctagaagtt	gaagcagaaa	ttgaccaagc	agattatagt	2400
gtattttact	aagccactct	ggaggaagaa	atggattcaa	aatcttttga	tgaaatggaa	2460
cagagcttac	ttattctttc	tgaaaccaag	gctkctctag	tgagcaccat	gagcctttgg	2520
aaacaacaga	tgtatacaat	agcaaagttt	catttcttta	ccttgaaacg	tgaaagtaaa	2580
tcagttagat	cagtgttgct	tctgctttta	atttttttca	cagttcagat	ttttatgttt	2640
ttggttcatc	actcttttaa	aaatgctgtg	gttcccatca	aacttgttcc	agacttatat	2700
tttctaaaac	ctggagacaa	accacataaa	tacaaaacaa	gtctgcttct	tcaaaattct	2760
gctgactcag	atatcagtga	tcttattagc	tttttcacaa	gccagaacat	aatgggtgacg	2820
atgattaatg	acagtgacta	tgtatccgtg	gtcctccata	gtgcggcctt	aaatgtgrtg	2880
cattcagaaa	aggactatgt	ttttgcagct	gttttcaaca	gtactatggg	ttattcttta	2940
cctatattag	tgaatatcat	tagtaactac	tacttttatt	atttaaatgt	gactgaaacc	3000
atccagatct	ggagtacccc	attctttcaa	gaaattactg	atatagtttt	taaaattgag	3060
ctgtattttc	aagcagcttt	gcttggaatc	attgttactg	caatgccacc	ttactttgcc	3120
atggaaaatg	cagagaatca	taagatcaaa	gcttayactc	aacttaaaact	ttcaggtctt	3180
ttgccatctg	catattggat	tggaacaagc	gttggtgata	tccccctatt	ttttatcatt	3240
cttatttttg	tgctaggaag	cttattggca	tttcattatg	gattatattt	ttatactgta	3300
aagttccttg	ctgtgggttt	ttgccttatt	ggttatgttc	catcagttat	tctgttcaact	3360
tatatgtgct	ctttcacctt	taagaaaatt	ttaaatacca	aagaattttg	gtcattttatc	3420
tattctgtgg	cagcgttggc	ttgtattgca	atcactgaaa	taactttctt	tatgggatac	3480
acaattgcaa	ctattcttca	ttatgccttt	tgtatcatca	ttccaatcta	tccacttcta	3540
ggttgccctg	tttctttcat	aaagatttct	tggaagaatg	tacgaaaaaa	tgtggacacc	3600
tataatccat	gggataggct	ttcagtagct	gttatatcgc	cttacctgca	gtgtgtactg	3660
tggatttttc	tcttacaata	ctatgagaaa	aaatatggag	gcagatcaat	aagaaaagat	3720
ccctttttca	gaaacctttc	aacgaagtct	aaaaatagga	agcttccaga	accaccagac	3780
aatgaggatg	aagatgaaga	tgtcaaagct	gaaagactaa	aggtcaaaga	gctgatgggt	3840
tgccagtggt	gtgaggagaa	accatccatt	atggctagca	atgtgcataa	agaatatgat	3900
gacaagaaag	attttcttct	ttcaagaaaa	gtaaagaaag	tggcaactaa	atacatctct	3960


```

ttctgtgtga aaaaaggaga gatccttagga ctattgggtc caaatgggtgc tggcaaaagc 4020
acaattatta atattctggt tggatgatatt gaaccaactt caggccaggt atttttagga 4080
gattattctt cagagacaag tgaagatgat gattcactga agtgtatggg ttactgtcct 4140
cagataaacc ctttgtggcc agatactaca ttgcaggaac attttgaaat ttatggagct 4200
gtcaaaggaa tgagtgaag tgacatgaaa gaagtcataa gtcgaataac acatgcactt 4260
gatttaaaag aacatcttca gaagactgta aagaactac ctgcaggaat caaacgaaag 4320
ttgtgttttg ctctaagtat gctaggggaat cctcagatta ctttgctaga tgaaccatct 4380
acaggatatg atcccaaagc caaacagcac atgtggcgag caattcgaac tgcatttaaa 4440
aacagaaagc gggctgctat tctgaccact cactatatgg aggaggcaga ggctgtctgt 4500
gatcgagtag ctatcatggt gtctgggcag ttaagatgta tcggaacagt acaacatcta 4560
aagagtaaat ttggaaaagg ctactttttg gaaattaaat tgaaggactg gatagaaaac 4620
ctagaagtag accgccttca aagagaaatt cagtatatatt tcccaaagtc aagccgtcag 4680
gaaagttttt cttctatttt ggcttataaa attcctaagg aagatgttca gtccttttca 4740
caatcttttt ttaagctgga agaaggtttt tgtagaactc actaa 4785

```

<210> 4

<211> 1594

<212> PRT

<213> homo sapiens

<400> 4

```

Met Ser Thr Ala Ile Arg Glu Val Gly Val Trp Arg Gln Thr Arg Thr
  1             5             10             15
Leu Leu Leu Lys Asn Tyr Leu Ile Lys Cys Arg Thr Lys Lys Ser Ser
  20             25             30
Val Gln Glu Ile Leu Phe Pro Leu Phe Phe Leu Phe Trp Leu Ile Leu
  35             40             45
Ile Ser Met Met His Pro Asn Lys Lys Tyr Glu Glu Val Pro Asn Ile
  50             55             60
Glu Leu Asn Pro Met Asp Lys Phe Thr Leu Ser Asn Leu Ile Leu Gly
  65             70             75             80
Tyr Thr Pro Val Thr Asn Ile Thr Ser Ser Ile Met Gln Lys Val Ser
  85             90             95
Thr Asp His Leu Pro Asp Val Ile Ile Thr Glu Glu Tyr Thr Asn Glu
 100             105             110
Lys Glu Met Leu Thr Ser Ser Leu Ser Lys Pro Ser Asn Phe Val Gly
 115             120             125
Val Val Phe Lys Asp Ser Met Ser Tyr Glu Leu Arg Phe Phe Pro Asp
 130             135             140
Met Ile Pro Val Ser Ser Ile Tyr Met Asp Ser Arg Ala Gly Cys Ser
 145             150             155             160
Lys Ser Cys Glu Ala Ala Gln Tyr Trp Ser Ser Gly Phe Thr Val Leu
 165             170             175
Gln Ala Ser Ile Asp Ala Ala Ile Ile Gln Leu Lys Thr Asn Val Ser
 180             185             190
Leu Trp Lys Glu Leu Glu Ser Thr Lys Ala Val Ile Met Gly Glu Thr
 195             200             205
Ala Val Val Glu Ile Asp Thr Phe Pro Arg Gly Val Ile Leu Ile Tyr
 210             215             220
Leu Val Ile Ala Phe Ser Pro Phe Gly Tyr Phe Leu Ala Ile His Ile
 225             230             235             240
Val Ala Glu Lys Glu Lys Lys Ile Lys Glu Phe Leu Lys Ile Met Gly
 245             250             255
Leu His Asp Thr Ala Phe Trp Leu Ser Trp Val Leu Leu Tyr Thr Ser
 260             265             270
Leu Ile Phe Leu Met Ser Leu Leu Met Ala Val Ile Ala Thr Ala Ser

```

275	280	285
Leu Leu Phe Pro Gln Ser Ser Ser Ile Val Ile Phe Leu Leu Phe Phe		
290	295	300
Leu Tyr Gly Leu Ser Ser Val Phe Phe Ala Leu Met Leu Thr Pro Leu		
305	310	315
Phe Lys Lys Ser Lys His Val Gly Ile Val Glu Phe Phe Val Thr Val		
325	330	335
Ala Phe Gly Phe Ile Gly Leu Met Ile Ile Leu Ile Glu Ser Phe Pro		
340	345	350
Lys Ser Leu Val Trp Leu Phe Ser Pro Phe Cys His Cys Thr Phe Val		
355	360	365
Ile Gly Ile Ala Gln Val Met His Leu Glu Asp Phe Asn Glu Gly Ala		
370	375	380
Ser Phe Ser Asn Leu Thr Ala Gly Pro Tyr Pro Leu Ile Ile Thr Ile		
385	390	395
Ile Met Leu Thr Leu Asn Ser Ile Phe Tyr Val Leu Leu Ala Val Tyr		
405	410	415
Leu Asp Gln Val Ile Pro Gly Glu Phe Gly Leu Arg Arg Ser Ser Leu		
420	425	430
Tyr Phe Leu Lys Pro Ser Tyr Trp Ser Lys Ser Lys Arg Asn Tyr Glu		
435	440	445
Glu Leu Ser Glu Gly Asn Val Asn Gly Asn Ile Ser Phe Ser Glu Ile		
450	455	460
Ile Glu Pro Val Ser Ser Glu Phe Val Gly Lys Glu Ala Ile Arg Ile		
465	470	475
Ser Gly Ile Gln Lys Thr Tyr Arg Lys Lys Gly Glu Asn Val Glu Ala		
485	490	495
Leu Arg Asn Leu Ser Phe Asp Ile Tyr Glu Gly Gln Ile Thr Ala Leu		
500	505	510
Leu Gly His Ser Gly Thr Gly Lys Ser Thr Leu Met Asn Ile Leu Cys		
515	520	525
Gly Leu Cys Pro Pro Ser Asp Gly Phe Ala Ser Ile Tyr Gly His Arg		
530	535	540
Val Ser Glu Ile Asp Glu Met Phe Glu Ala Arg Lys Met Ile Gly Ile		
545	550	555
Cys Pro Gln Leu Asp Ile His Phe Asp Val Leu Thr Val Glu Glu Asn		
565	570	575
Leu Ser Ile Leu Ala Ser Ile Lys Gly Ile Pro Ala Thr Asn Ile Ile		
580	585	590
Gln Glu Val Gln Lys Val Leu Leu Asp Leu Asp Met Gln Thr Ile Lys		
595	600	605
Asp Asn Gln Ala Lys Lys Leu Ser Gly Gly Gln Lys Arg Lys Leu Ser		
610	615	620
Leu Gly Ile Ala Val Leu Gly Asn Pro Lys Ile Leu Leu Leu Asp Glu		
625	630	635
Pro Thr Ala Gly Met Asp Pro Cys Ser Arg His Ile Val Trp Asn Leu		
645	650	655
Leu Lys Tyr Arg Lys Ala Asn Arg Val Thr Val Phe Ser Thr His Phe		
660	665	670
Met Asp Glu Ala Asp Ile Leu Ala Asp Arg Lys Ala Val Ile Ser Gln		
675	680	685
Gly Met Leu Lys Cys Val Gly Ser Ser Met Phe Leu Lys Ser Lys Trp		
690	695	700
Gly Ile Gly Tyr Arg Leu Ser Met Tyr Ile Asp Lys Tyr Cys Ala Thr		
705	710	715
Glu Ser Leu Ser Ser Leu Val Lys Gln His Ile Pro Gly Ala Thr Leu		

1170	1175	1180
Ser Phe Ile Lys Ile	Ser Trp Lys Asn Val	Arg Lys Asn Val Asp Thr
1185	1190	1195 1200
Tyr Asn Pro Trp Asp	Arg Leu Ser Val	Ala Val Ile Ser Pro Tyr Leu
1205	1210	1215
Gln Cys Val Leu Trp	Ile Phe Leu Leu Gln Tyr Tyr	Glu Lys Lys Tyr
1220	1225	1230
Gly Gly Arg Ser Ile	Arg Lys Asp Pro Phe Phe	Arg Asn Leu Ser Thr
1235	1240	1245
Lys Ser Lys Asn Arg	Lys Leu Pro Glu Pro	Pro Asp Asn Glu Asp Glu
1250	1255	1260
Asp Glu Asp Val Lys	Ala Glu Arg Leu Lys	Val Lys Glu Leu Met Gly
1265	1270	1275 1280
Cys Gln Cys Cys Glu	Glu Lys Pro Ser Ile	Met Val Ser Asn Leu His
1285	1290	1295
Lys Glu Tyr Asp Asp	Lys Lys Asp Phe Leu Leu	Ser Arg Lys Val Lys
1300	1305	1310
Lys Val Ala Thr Lys	Tyr Ile Ser Phe Cys Val	Lys Lys Gly Glu Ile
1315	1320	1325
Leu Gly Leu Leu Gly	Pro Asn Gly Ala Gly	Lys Ser Thr Ile Ile Asn
1330	1335	1340
Ile Leu Val Gly Asp	Ile Glu Pro Thr Ser	Gly Gln Val Phe Leu Gly
1345	1350	1355 1360
Asp Tyr Ser Ser Glu	Thr Ser Glu Asp Asp	Asp Ser Leu Lys Cys Met
1365	1370	1375
Gly Tyr Cys Pro Gln	Ile Asn Pro Leu Trp	Pro Asp Thr Thr Leu Gln
1380	1385	1390
Glu His Phe Glu Ile	Tyr Gly Ala Val Lys	Gly Met Ser Ala Ser Asp
1395	1400	1405
Met Lys Glu Val Ile	Ser Arg Ile Thr His	Ala Leu Asp Leu Lys Glu
1410	1415	1420
His Leu Gln Lys Thr	Val Lys Lys Leu Pro	Ala Gly Ile Lys Arg Lys
1425	1430	1435 1440
Leu Cys Phe Ala Leu	Ser Met Leu Gly Asn	Pro Gln Ile Thr Leu Leu
1445	1450	1455
Asp Glu Pro Ser Thr	Gly Met Asp Pro Lys	Ala Lys Gln His Met Trp
1460	1465	1470
Arg Ala Ile Arg Thr	Ala Phe Lys Asn Arg	Lys Arg Ala Ala Ile Leu
1475	1480	1485
Thr Thr His Tyr Met	Glu Glu Ala Glu Ala	Val Cys Asp Arg Val Ala
1490	1495	1500
Ile Met Val Ser Gly	Gln Leu Arg Cys Ile	Gly Thr Val Gln His Leu
1505	1510	1515 1520
Lys Ser Lys Phe Gly	Lys Gly Tyr Phe Leu	Glu Ile Lys Leu Lys Asp
1525	1530	1535
Trp Ile Glu Asn Leu	Glu Val Asp Arg Leu	Gln Arg Glu Ile Gln Tyr
1540	1545	1550
Ile Phe Pro Asn Ala	Ser Arg Gln Glu Ser	Phe Ser Ser Ile Leu Ala
1555	1560	1565
Tyr Lys Ile Pro Lys	Glu Asp Val Gln Ser	Leu Ser Gln Ser Phe Phe
1570	1575	1580
Lys Leu Glu Glu Gly	Phe Cys Arg Thr His	
1585	1590	

<210> 5

<211> 5262

<212> DNA

<213> homo sapiens

<400> 5

actgttgata	tgggtggtatt	tcaaattctg	gtctacccta	tttcacatgc	cttgtttact	60
tttcagagct	gacagattgc	tgctccatgc	attctgtcca	gtttcctaag	agagacagct	120
tggagtatgc	ttaatccatc	ttacctggga	ctgaaacagc	tgcttatttt	gccgttaaaa	180
attacatgca	gtttactgcg	tggctccggg	tttgtttgtt	tgtttttcct	ctttaatagg	240
tttattcaga	aaacatgtcc	actgcaatta	gggaggtagg	agtttgagga	cagaccagaa	300
cacttctact	gaagaattac	ttaattaaat	gcagaaccaa	aaagagtagt	gttcaggaaa	360
ttctttttcc	actatTTTTT	ttatTTTgtt	taatattaat	tagcatgatg	catccaaata	420
agaaatatga	agaagtgcct	aatatagaac	tcaatcctat	ggacaagttt	actctttcta	480
atctaattct	tggatatact	ccagtgacta	atattacaag	cagcatcatg	cagaaagtgt	540
ctactgatca	tctacctgat	gtcataatta	ctgaagaata	tacaaatgaa	aaagaaatgt	600
taacatccag	tctctctaag	ccgagcaact	ttgtagggtg	ggttttcaaa	gactccatgt	660
cctatgaact	tcgttttttt	cctgatatga	ttccagtatc	ttctatttat	atggattcaa	720
gagctggctg	ttcaaaatca	tgtgaggctg	ctcagtactg	gtcctcaggt	ttcacagttt	780
tacaagcatc	catagatgct	gccattatac	agttgaagac	caatgtttct	ctttggaagg	840
agctggagtc	aactaaagct	gttattatgg	gagaaactgc	tgttgtagaa	atagatacct	900
ttccccgagg	agtaatttta	atatacctag	ttatagcatt	ttcacctttt	ggatactttt	960
tggcaattca	tatcgtagca	gaaaaagaaa	aaaaaataaa	agaattttta	aagataatgg	1020
gacttcatga	tactgccttt	tggctttcct	gggttcttct	atawacaagt	ttaatttttc	1080
ttatgtccct	tcttatggca	gtcattgcga	cagcttcttt	gttatttcct	caaagtagca	1140
gcattgtgat	atttctgctt	tttttccctt	atggattatc	atctgtattt	tttgctttaa	1200
tgctgacacc	tctttttaaa	aaatcaaaac	atgtgggaat	agttgaattt	tttgttactg	1260
tggcttttgg	atttattggc	cttatgataa	tcctcataga	aagttttccc	aaatcgttag	1320
tgtggctttt	cagtcctttc	tgtcactgta	cttttgtgat	tggtattgca	caggtcatgc	1380
atttgaaga	ttttaatgaa	ggtgcttcat	tttcaaattt	gactgcaggc	ccatatacct	1440
taattattac	aattatcatg	ctcacactta	atagtatatt	ctatgtcctc	ttggctgtct	1500
atcttgatca	agtcattcca	ggggaatttg	gcttacggag	atcatcttta	tattttctga	1560
agccttcata	ttggtcaaa	agcaaaagaa	attatgagga	gttatcagag	ggcaatgtta	1620
atggaaatat	tagtttttagt	gaaattattg	agccagtttc	ttcagaattt	gtaggaaaag	1680
aagccataag	aattagtggg	attcagaaga	catacagaaa	gaagggtgaa	aatgtggagg	1740
ctttgagaaa	tttgtcattt	gacatatatg	agggtcagat	tactgcctta	cttggccaca	1800
gtggaacagg	aaagagtaca	ttgatgaata	ttctttgtgg	actctgccca	ccttctgatg	1860
ggtttgcata	tatatatgga	cacagagtct	cagaaataga	tgaaatgttt	gaagcaagaa	1920
aaatgattgg	catttgtcca	cagttagata	tacactttga	tgttttgaca	gtagaagaaa	1980
atttatcaat	tttggcttca	atcaaaggga	taccagccac	caatataata	caagaagtgc	2040
agaaggtttt	actagattta	gacatgcaga	ctatcaaaga	taaccaagct	aaaaaattaa	2100
gtggtgggtca	aaaaagaaag	ctgtcattag	gaattgtctg	tcttgggaac	ccaaagatac	2160
tgctgctaga	tgaaccaaca	gctggaatgg	accctgttcc	tcgacataat	gtatggaatc	2220
ttttaaaata	cagaaaagcc	aatcgggtga	cagtgttcag	tactcatttc	atggatgaag	2280
ctgacattct	tgcagatagg	aaagctgtga	tatcacaagg	aatgctgaaa	tgtgttggtt	2340
cttcaatgtt	cctcaaaagt	aaatggggga	tcggctaccg	cctgagcatg	tacatagaca	2400
aatattgtgc	cacagaatct	ctttcttcac	tggttaaaca	acatatacct	ggagctactt	2460
tattacaaca	gaatgaccaa	caacttgtgt	atagcttgcc	tttcaaggac	atggacaaat	2520
tttcaggttt	gttttctgcc	ctagacagtc	attcaaat	gggtggcatt	tcttatgggg	2580
tttccatgac	gactttggaa	gacgtatttt	taaagctaga	agttgaagca	gaaattgacc	2640
aagcagatta	tagtgtat	actcagcagc	cactggagga	agaaatggat	tcaaaatctt	2700
ttgatgaaat	ggaacagagc	ttacttattc	tttctgaaac	caaggctkct	ctagtggagca	2760
ccatgagcct	ttggaacaaa	cagatgtata	caatagcaaa	gtttcatttc	tttaccttga	2820
aacgtgaaag	taaatacagt	agatcagtg	tgcttctgct	tttaattttt	ttcacagttc	2880
agatttttat	gtttttgggt	catcactctt	ttaaaaatgc	tgtggttccc	atcaaacttg	2940
ttccagactt	atattttcta	aaacctggag	acaaaccaca	taaatacaaa	acaagtctgc	3000
ttcttcaaaa	ttctgctgac	tcagatatca	gtgatcttat	tagctttttc	acaagccaga	3060
acataatggg	gacgatgatt	aatgacagtg	actatgtatc	cgtggctccc	catagtgcgg	3120

ctttaaatgt	grtgcattca	gaaaaggact	atgtttttgc	agctgttttc	aacagtacta	3180
tgggtttattc	tttacctata	ttagtgaata	tcattagtaa	ctactatcct	tatcatttaa	3240
atgtgactga	aaccatccag	atctggagta	ccccattcct	tcaagaaatt	actgatatag	3300
tttttaaaat	tgagctgtat	tttcaagcag	ctttgcttgg	aatcattggt	actgcaatgc	3360
caccttactt	tgccatggaa	aatgcagaga	atcataagat	caaagcttay	actcaactta	3420
aactttcagg	tcttttgcca	tctgcatatt	ggattggaca	agctgttggt	gatatcccct	3480
tattttttat	cattcttatt	ttgatgctag	gaagcttatt	ggcatttcat	tatggattat	3540
attttttatac	tgtaaagtcc	cttgctgtgg	ttttttgcct	tattggttat	gttccatcag	3600
ttattctgtt	cacttatatt	gcttctttca	cctttaagaa	aatttttaa	accaaagaat	3660
tttggtcatt	tatctattct	gtggcagcgt	tggcttgtat	tgcaatcact	gaaataactt	3720
tctttatggg	atacacaatt	gcaactatcc	ttcattatgc	cttttgtatc	atcattccaa	3780
tctatccact	tctaggttgc	ctgatttctt	tcataaagat	ttcttggag	aatgtacgaa	3840
aaaatgtgga	cacctataat	ccatgggata	ggcttccagt	agctgttata	tcgccttacc	3900
tgcagtgtgt	actgtggatt	ttcctcttac	aatactatga	gaaaaaatat	ggaggcagat	3960
caataagaaa	agatcccctt	ttcagaaacc	tttcaacgaa	gtctaaaaat	aggaagcttc	4020
cagaaccacc	agacaatgag	gatgaagatg	aagatgtcaa	agctgaaaga	ctaaagggtca	4080
aagagctgat	gggttgccag	tgttgtgagg	agaaaccatc	cattatgggtc	agcaatttgc	4140
ataaagaata	tgatgacaag	aaagattttc	ttctttcaag	aaaagtaaag	aaagtggcaa	4200
ctaaatacat	ctctttctgt	gtgaaaaaag	gagagatcct	aggactattg	ggtccaaatg	4260
gtgctggcaa	aagcacaatt	attaatatcc	tgggttggtga	tattgaacca	acttcaggcc	4320
aggatatttt	aggagattat	tcttcagaga	caagtgaaga	tgatgattca	ctgaagtgtg	4380
tgggttactg	tcctcagata	aaccctttgt	ggccagatac	tacattgcag	gaacattttg	4440
aaatttatgg	agctgtcaaa	ggaatgagtg	caagtgcacat	gaaagaagtc	ataagtcgaa	4500
taacacatgc	acttgattta	aaagaacatc	ttcagaagac	tgtaaagaaa	ctacctgcag	4560
gaatcaaacy	aaagtgtgtg	tttgctctaa	gtatgctagg	gaatcctcag	attactttgc	4620
tagatgaacc	atctacaggt	atggatccca	aagccaaaca	gcacatgtgg	cgagcaattc	4680
gaactgcatt	taaaaacaga	aagcgggctg	ctattctgac	cactcactat	atggaggagg	4740
cagaggctgt	ctgtgatcga	gtagctatca	tgggtgtctgg	gcagttaaga	tgatcggaa	4800
cagtacaaca	tctaaagagt	aaatttggaa	aaggctactt	tttggaaatt	aaattgaagg	4860
actgगतaga	aaacctagaa	gtagaccgcc	ttcaaagaga	aattcagtat	attttcccaa	4920
atgcaagccg	tcaggaaaagt	ttttcttcta	ttttggctta	taaaattcct	aaggaagatg	4980
ttcagtcctt	ttcacaatct	ttttttaagc	tgggaagaagc	taaacatgct	tttgccattg	5040
aagaatatag	cttttctcaa	gcaacattgg	aacaggtttt	tgtagaactc	actaaagaac	5100
aagaggagga	agataatagt	tgtggaactt	taaacagcac	actttgggtg	gaacgaacac	5160
aagaagatag	agtagtatct	tgaatttgta	ttgttcgggc	tgcttactgg	gacttctttc	5220
tttttcactt	aattttaact	ttgggtttaa	aagtttttat	tg		5262